

PCT

ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international



DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁵ : C12N 15/77, 15/62, 1/21 C12N 15/90, C07K 13/00, 15/00 A61K 39/40		A1	(11) Numéro de publication internationale: WO 93/03158 (43) Date de publication internationale: 18 février 1993 (18.02.93)																																																								
(21) Numéro de la demande internationale: PCT/FR92/00744 (22) Date de dépôt international: 29 juillet 1992 (29.07.92)		(74) Mandataire: WARCOIN, Jacques; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).																																																									
(30) Données relatives à la priorité: 91/09652 30 juillet 1991 (30.07.91) 91/09870 2 août 1991 (02.08.91)		FR FR	(81) Etats désignés: JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).																																																								
(71) Déposant (pour tous les Etats désignés sauf US): ORSAN [FR/FR]; 16, rue Ballu, F-75009 Paris (FR).		Publiée <i>Avec rapport de recherche internationale.</i>																																																									
(72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): JOLIFF, Gwennaël [FR/FR]; 48, rue Truffaut, F-75017 Paris (FR). GUYONVARCH, Armel [FR/FR]; 21, avenue Flouquet, F-94240 L'Hay-les-Roses (FR). RELANO, Purification [FR/FR]; 12, rue Pierrelais, F-92240 Fontenay-aux-Roses (FR). LEBLON, Gérard [FR/FR]; 5, allée des Bathes, F-91940 Les Ulis (FR). DUCHIRON, Francis [FR/FR]; Résidence La Fontaine-au-Bois, 25, rue Mermoz, F-77210 Avon (FR). RENAUD, Michel [FR/FR]; 23, rue des Causses, F-91940 Les Ulis (FR).																																																											
<p>(54) Titre: SYSTEM FOR PROTEIN EXPRESSION AND SECRETION ESPECIALLY IN CORYNEBACTERIA</p> <p>(54) Titre: SYSTEME D'EXPRESSION ET DE SECRETION DE PROTEINES UTILISABLES EN PARTICULIER CHEZ LES CORYNEBACTERIES</p> <p>(57) Abstract</p> <p>System for the expression and secretion of an amino acid, a polypeptide or a protein determined by a corynebacteria strain. The system is characterized in that the sequence coding for said amino acid, polypeptide or protein is located in a chromosomal or plasmid DNA region where said sequence is transcribed towards the 5' end with at least one part of the sequence coding for the signal sequence of protein PS1 or PS2, said part secreting said protein after translation upon incorporation of the system into said corynebacteria strain.</p> <p>(57) Abrégé</p> <p>La présente invention concerne un système d'expression et de sécrétion d'un amino-acide, polypeptide ou protéine déterminé par une souche de corynébactéries, caractérisé en ce que la séquence qui code pour ledit amino-acide, polypeptide ou ladite protéine est située dans une région d'ADN chromosomique ou plasmidique où ladite séquence est transcrit avec vers l'extrémité 5' au moins une partie de la séquence codant pour la séquence signal de la protéine PS1 ou PS2, ladite partie assurant la sécrétion de ladite protéine après traduction lorsque le système est incorporé dans ladite souche de corynébactéries.</p> <table border="1"><thead><tr><th rowspan="2">Précédent - PRECURSOR</th><th rowspan="2">Mature - MATURED</th><th rowspan="2">Sécrétion - SECRETION</th><th rowspan="2">EXPECTED PROTEIN SIZE: (D_n)</th><th colspan="2">IMMUNOLOGICAL PROTEIN DETECTION</th></tr><tr><th>TAILLES ATTENDUES DES PROTEINES EN D_n</th><th>DÉTECTION IMMUNOLOGIQUE DES PROTÉINES</th></tr></thead><tbody><tr><td>pCGL616</td><td>NcoI</td><td>Poul Nasl Nhel</td><td>70 887 66 477 67 000</td><td>+</td><td>+++</td></tr><tr><td>pCGL1036</td><td></td><td>NarI</td><td>65 055 60 545 58 000</td><td>+</td><td>+</td></tr><tr><td>pCGL1037</td><td></td><td>SnaBI BspEI</td><td>58 592 54 082 56 000</td><td>+</td><td>++</td></tr><tr><td>pCGL1038</td><td></td><td></td><td>69 955 55 444 60 000</td><td>+</td><td>++</td></tr><tr><td>pCGL1039</td><td></td><td></td><td>40 189 35 679 35 000</td><td>+</td><td>+</td></tr><tr><td>pCGL1040</td><td></td><td></td><td>29 854 25 343 32 000</td><td>+</td><td>+</td></tr><tr><td>pCGL1041</td><td></td><td></td><td>28 694 24 183 23 000</td><td>+</td><td>+</td></tr><tr><td>pCGL1042</td><td></td><td></td><td>13 498 8 987</td><td>-</td><td>-</td></tr></tbody></table> <p>Légende: ■ Séquence signalée putative. - PUTATIVE SIGNAL SEQUENCE LEGEND: ■ protéine mature. - MATURED PROTEIN ■ Terminateur. - TERMINATOR</p>				Précédent - PRECURSOR	Mature - MATURED	Sécrétion - SECRETION	EXPECTED PROTEIN SIZE: (D _n)	IMMUNOLOGICAL PROTEIN DETECTION		TAILLES ATTENDUES DES PROTEINES EN D _n	DÉTECTION IMMUNOLOGIQUE DES PROTÉINES	pCGL616	NcoI	Poul Nasl Nhel	70 887 66 477 67 000	+	+++	pCGL1036		NarI	65 055 60 545 58 000	+	+	pCGL1037		SnaBI BspEI	58 592 54 082 56 000	+	++	pCGL1038			69 955 55 444 60 000	+	++	pCGL1039			40 189 35 679 35 000	+	+	pCGL1040			29 854 25 343 32 000	+	+	pCGL1041			28 694 24 183 23 000	+	+	pCGL1042			13 498 8 987	-	-
Précédent - PRECURSOR	Mature - MATURED	Sécrétion - SECRETION	EXPECTED PROTEIN SIZE: (D _n)					IMMUNOLOGICAL PROTEIN DETECTION																																																			
				TAILLES ATTENDUES DES PROTEINES EN D _n	DÉTECTION IMMUNOLOGIQUE DES PROTÉINES																																																						
pCGL616	NcoI	Poul Nasl Nhel	70 887 66 477 67 000	+	+++																																																						
pCGL1036		NarI	65 055 60 545 58 000	+	+																																																						
pCGL1037		SnaBI BspEI	58 592 54 082 56 000	+	++																																																						
pCGL1038			69 955 55 444 60 000	+	++																																																						
pCGL1039			40 189 35 679 35 000	+	+																																																						
pCGL1040			29 854 25 343 32 000	+	+																																																						
pCGL1041			28 694 24 183 23 000	+	+																																																						
pCGL1042			13 498 8 987	-	-																																																						

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures
publiant des demandes internationales en vertu du PCT.

AT	Autriche	FI	Finlande	MN	Mongolie
AU	Australie	FR	France	MR	Mauritanie
BB	Barbade	CA	Gabon	MW	Malawi
BE	Belgique	GB	Royaume-Uni	NL	Pays-Bas
BF	Burkina Faso	GN	Guinée	NO	Norvège
BG	Bulgarie	GR	Grèce	NZ	Nouvelle-Zélande
BJ	Bénin	HU	Hongrie	PL	Pologne
BR	Brazil	IE	Irlande	PT	Portugal
CA	Canada	IT	Italie	RO	Roumanie
CF	République Centrafricaine	JP	Japon	RU	Fédération de Russie
CG	Congo	KP	République populaire démocratique de Corée	SD	Soudan
CH	Suisse	KR	République du Coréa	SE	Suède
CI	Côte d'Ivoire	LJ	Liechtenstein	SK	République slovaque
CM	Cameroon	LK	Sri Lanka	SN	Sénégal
CS	Tchécoslovaquie	LU	Luxembourg	SU	Union soviétique
CZ	République tchèque	MC	Monaco	TD	Tchad
DE	Allemagne	MG	Madagascar	TG	Togo
DK	Danemark	ML	Mali	UA	Ukraine
ES	Espagne			US	Etats-Unis d'Amérique

SYSTEME D'EXPRESSION ET DE SECRETION DE PROTEINES
UTILISABLES EN PARTICULIER CHEZ LES CORYNEBACTERIES

La présente invention concerne notamment des systèmes d'expression et de sécrétion de protéines utilisables, en particulier, chez les corynénactéries, ainsi qu'un procédé mettant en œuvre ces systèmes, ainsi que de nouvelles protéines liées à ces systèmes d'expression.

Les corynénactéries constituent un groupe de bactéries gram positif, de morphologie irrégulière, représenté par une grande variété de souches.

10 En dépit du fait que les cellules gram positif ont une structure simple facilitant la sécrétion de protéines dans le milieu extérieur, la sécrétion des protéines par les corynénactéries a été très peu étudiée jusqu'ici. Seules la toxine diphérique secrétée par certaines souches de Corynebacterium diphtheriae infectées par des phages lysogéniques tox+ (Smith 1980 : J. Bacteriol. 141 pp 1142 ; Smith et al. 1980 : J. Bacteriol. 141 pp 184 ; Greenfield et al. 1983 : PNAS USA 80 pp 6853) et l'étude de la séquence nucléotidique d'un gène impliqué dans la sécrétion d'une DNase par Corynebacterium glutamicum (W. Liebl et A.S. Sinskey 1986 : Genetics and Biotechnology of Bacteria, Vol. 2 pp 383-388) ont été rapportées.

20 Le brevet américain US 4 965 197 décrit un système d'expression et de sécrétion utilisable chez Corynebacterium sur la base de la DNase décrite précédemment ; néanmoins il semble que cette protéine ne soit pas majoritaire et que dans ces conditions, le système de sécrétion correspondant soit peu intéressant.

25 C'est pourquoi l'invention concerne un système d'expression et de sécrétion dans une bactérie de type corynénactérie comportant les éléments de sécrétion de deux protéines se trouvant majoritairement dans le surnageant de culture de certaines corynénactéries.

Il s'agit notamment d'un système d'expression/sécrétion d'un 30 amino-acide, polypeptide ou protéine déterminé par une souche de corynénactérie, caractérisé en ce que la séquence qui code pour ledit amino-acide, polypeptide ou ladite protéine est située dans une région d'ADN chromosomique ou plasmidique où ladite séquence est transcrive avec vers l'extrémité 5' au moins une partie de la séquence codant pour la 35 séquence signal de la protéine PS1 ou PS2, ladite partie assurant la

sécrétion dudit amino-acide, polypeptide ou de ladite protéine après traduction lorsque le système est incorporé dans ladite souche de corynébactérie.

Plus particulièrement, la présente invention concerne un système d'expression et de sécrétion dans une corynébactérie comprenant :

5 - une souche de corynébactérie,

- une cassette de sécrétion contenant une première séquence d'ADN fonctionnelle pour l'expression dans ladite souche de corynébactérie, une seconde séquence d'ADN qui code pour un aminoacide, un polypeptide et/ou

10 une protéine, et une troisième séquence d'ADN insérée entre lesdites première et seconde séquences d'ADN qui codent pour des éléments d'une protéine choisie parmi PS1 ou PS2, qui assurent la sécrétion desdits amino-acides, polypeptides et/ou protéines par ladite souche de corynébactérie.

15 Il doit tout d'abord être compris que dans le cadre de la présente invention la terminologie "corynébactérie" désigne non seulement les souches du genre Corynebacterium mais également les bactéries apparentées, telles que Brevibacterium.

20 Le système d'expression de la présente invention peut se trouver sur un plasmide autoréplicatif dans corynébactérie et dans ce cas, le plasmide comportera une origine de réPLICATION fonctionnelle dans la souche, par exemple dans Corynebacterium une origine de réPLICATION PBL 1, mais peut également être portée par un plasmide non réplicatif destiné notamment à l'intégration chromosomique, dans ce cas le plasmide 25 comportera des éléments permettant la recombinaison et l'intégration chromosomique ; dans le cas de l'intégration, le système d'expression se trouvera finalement dans le chromosome de la bactérie en cause.

En particulier, dans le cas de l'intégration chromosomique, on 30 a pu démontrer que l'insertion d'une séquence d'ADN hétérologue dans le gène csp1 codant pour PS1 ou csp2 codant pour PS2 n'affectait pas la croissance de la souche correspondante. Dans ces conditions, il est possible d'intégrer une séquence codant pour un amino-acide, un polypeptide ou une protéine en phase dans csp1 ou csp2 afin d'obtenir l'expression/sécrétion 35 des produits d'expression de la séquence codante insérée.

Parmi les séquences d'ADN fonctionnelles pour l'expression dans la souche de corynébactérie il faut citer aussi bien des éléments

d'expression homologues que des éléments d'expression hétérologues, c'est-à-dire qu'il peut s'agir d'éléments qui existent déjà dans la bactérie hôte ou bien qui au contraire, proviennent d'une bactérie différente.

5 Ces éléments d'expression comporteront essentiellement un promoteur et un site de fixation des ribosomes, mais il est possible également de prévoir d'autres éléments notamment du type éléments de régulation de l'expression.

10 Parmi les éléments d'expression utilisables dans les corynèbactéries, on utilisera plus particulièrement le promoteur Ptac qui est un promoteur fort, hybride trp/lac inductible par IPTG et qui se révèle fonctionnel chez les corynèbactéries comme chez E. coli ; mais il est possible d'utiliser d'autres promoteurs, par exemple, comme cela sera décrit ci-après, des promoteurs ou d'autres éléments d'expression de gène de structure de corynèbactérie, par exemple le promoteur de la gdhA. Il est possible également de prévoir d'utiliser, par exemple, les éléments d'expression, notamment le promoteur de l'une des protéines PS1 et/ou PS2 tels qu'ils sont identifiés dans le cadre de la présente invention.

15 Les éléments d'expression peuvent également comporter des séquences d'ADN assurant la régulation de l'expression des gènes en aval.

20 Parmi les éléments assurant une bonne expression on peut prévoir de pouvoir placer à la fin de la séquence codante un élément d'arrêt de traduction, sous forme d'un ou plusieurs codons stop, ou un élément d'arrêt de la transcription.

25 Parmi les éléments assurant la sécrétion il faut citer, comme cela a été indiqué précédemment, tout ou partie de la séquence signal de l'une des protéines PS1 ou PS2 ainsi que les équivalents de ces séquences, sans modification ou perte de la propriété de sécrétion.

30 En effet, il est certain qu'il est possible d'apporter grâce à des techniques connues telles que des mutations ponctuelles des modifications mineures aux séquences de sécrétion, tout en conservant le même type de propriété de sécrétion, c'est pourquoi la présente invention entend également couvrir ces séquences équivalentes.

35 Le système d'expression de la présente invention peut enfin comporter d'autres éléments, notamment des éléments tels que des terminateurs de transcription, par exemple le terminateur des protéines PS1 et/ou PS2, ou de la gdhA.

Dans certains cas, il peut être intéressant d'ajouter aux séquences d'expression et de sécrétion, toute ou partie de la protéine PS1 afin d'obtenir des protéines fusionnées, dont la sécrétion et le taux d'expression peuvent dans ces conditions être améliorés.

5 Les systèmes d'expression selon l'invention, peuvent comporter des éléments hétérologues qui permettent la construction dans des bactéries différentes des corynébactéries par exemple comme cela a été dit précédemment une origine de réPLICATION fonctionnelle dans E. coli mais également d'autres éléments comme un gène marqueur qui peuvent faciliter 10 le transfert dans Corynebacterium.

Le gène marqueur peut être, bien entendu, de type très varié pour autant qu'il soit fonctionnel chez corynébactérie, il pourrait s'agir d'un gène de sélection positif ou négatif tel qu'une résistance spécifique, néanmoins dans l'état actuel des recherches ces gènes ne sont pas aisément 15 disponibles. On utilisera donc de préférence le gène celA de la cellulase de Clostridium thermocellum (celA) qui confère le phénotype CMC⁺, mais il est possible d'utiliser d'autres gènes marqueurs, notamment lacZ d'E. coli.

Dans le cas où le gène marqueur est celA, les bactéries transformées sont sélectionnées pour le caractère CMC⁺ après insertion de 20 la séquence codante dans un site de restriction approprié comme BstXI.

Dans le procédé selon l'invention, on prévoira, de préférence, que le gène marqueur puisse être facilement éliminé après vérification de la construction notamment en plaçant des sites de restriction entre le gène 25 marqueur et la séquence codante.

La séquence codante pourra être naturelle, synthétique ou mixte.

Le système d'expression et de sécrétion de la présente invention est bien entendu destiné plus particulièrement à assurer la production de produits d'intérêt industriel, c'est pourquoi les séquences 30 codantes coderont plus particulièrement pour un peptide, un polypeptide, ou une protéine d'intérêt industriel mais il pourra également s'agir d'une séquence codant non pas directement pour une protéine d'intérêt industriel,

mais pour une protéine pouvant intervenir dans la maturation et/ou l'élaboration d'un aminoacide, d'un polypeptide, ou d'une protéine présentant un intérêt industriel.

5 Les procédés selon l'invention sont plus particulièrement destinés à l'expression de séquences d'acides aminés, notamment des séquences répétitives, il s'agit donc principalement de séquences synthétiques.

Cette seconde séquence de DNA codant pour ces différents produits pourra également comporter certains éléments destinés à assurer la maturation du produit sécrété.

10 Dans le cas d'une séquence synthétique, le choix de la séquence codante permet la constitution :

- d'une séquence d'acides aminés quelconque ;
- d'une séquence d'acides aminés répétitive à n répétitions du type $(aa_1 \dots aa_x)_n$;

15 - d'une séquence répétitive contenant en position COOH-terminale aa_x un acide aminé chargé positivement ou négativement. Cet acide aminé peut permettre d'améliorer l'expression génétique mais permet avantageusement

(i) d'isoler le polypeptide du fait de son caractère ionique marqué ;
20 (ii) de cliver par protéase spécifique le polypeptide en unité $(aa_1 \dots aa_x)$;
(iii) d'enlever si nécessaire l'acide aminé terminal aa_x par des carboxy peptidases spécifiques ;

25 - d'une séquence répétitive contenant en portion NH_2 ou COOH terminale aa_1 ou aa_x un acide aminé conférant un avantage désiré.

Dans les exemples, la séquence exprimée code pour un polypeptide de structure $(ala-gln)_{20}$ et $(ala-gln-lys)_{10}$. La séquence ala-gln ou ala-gln-lys peut être libérée par traitement enzymatique ultérieur. On peut prévoir d'autres polymères de ce type tels que Ala-Gln-Tyr ou Ala-Gln-Met qui peuvent être libérés par traitement enzymatique ou chimique.

30 Le choix des codons de la séquence codante peut influencer l'expression dans les corynébactéries, il convient de préférence de prévoir une séquence ayant une richesse en GC de l'ordre de 50 à 60 %.

Dans le cas de l'exemple la séquence codant pour (AQ)₂₀ est GCX CAG avec X = A ou T ou C ou G ; en effet si aucun des codons n'est préféré pour l'alanine, par contre le codon CAG est très nettement préféré pour la glutamine. Dans ce cas, le pourcentage de GC est de l'ordre de 5 75 %, ce qui peut constituer une limitation, c'est pourquoi on envisage l'utilisation de polymères comportant 3 amino-acides dont le troisième est riche en A et T pour ramener le pourcentage à 55 %.

Tyr, Lys et Met possèdent deux A ou T dans les deux premières bases de leur codon et permettent donc de faire descendre le pourcentage en GC de 75 % à environ 60 %, ce qui devient plus proche du pourcentage en GC trouvé chez les corynèbactéries. D'autres part, bien sûr, l'intérêt industriel pour ces deux acides aminés en position COOH-terminal de la glutamine (Q) a été considéré et existe.

La présente invention concerne également les souches de corynèbactéries comportant un système d'expression et de sécrétion tel qu'il a été décrit précédemment, et plus particulièrement lorsque la souche est une Brevibacterium notamment une souche de Brevibacterium lactofermentum.

Enfin, la présente invention concerne un procédé d'obtention d'aminoacides, de polypeptides, ou de protéines caractérisé en ce que l'on cultive dans un milieu de culture une souche de corynèbactérie transformée telle que cela a été décrit précédemment, dans laquelle la seconde séquence de DNA code pour lesdits aminoacides ledit polypeptide, et/ou ladite protéine et en ce que l'on sépare éventuellement après culture ledit produit du milieu de culture. En effet, grâce à ce procédé, le produit intéressant a été sécrété et se trouve donc dans le milieu de culture duquel il peut être isolé par des procédés connus, qu'il s'agisse de techniques de séparation telles que la chromatographie, la précipitation sélective par exemple, celle-ci devant être adaptée évidemment à la nature de la molécule produite.

Il est également possible de prévoir la séparation du concentrat bactérien puis la séparation de la protéine déterminée, fusionnée ou non avec PS1 ou PS2 à partir de ce concentrat, par exemple par utilisation d'un agent tensio-actif. En effet, PS1 et PS2 étant des

protéines pariétales, une partie des protéines sécrétées avec ce système restent ancrées dans la paroi, ce qui peut faciliter leur séparation car avec certains détergents la bactérie n'est pas lysée.

La transformation des corynélactéries par des plasmides est réalisée de préférence par électroporation (Bonamy C., Guyonvach A., Reyes O., David F. and Leblon G. (1990) FEMS Microbiology Letters 66 : 263-270) ou par tout autre procédé approprié.

Les conditions de fermentation permettant la préparation d'amino acides, peptides et/ou protéines dépendent évidemment du type de produit obtenu ainsi que de la souche spécifique mise en oeuvre, il s'agit là d'éléments qui doivent être déterminés spécifiquement pour chaque souche en fonction des connaissances de l'homme de métier.

La présente invention concerne également des systèmes d'expression comportant tout ou partie des signaux d'expression de csp1, csp2 et de la gdhA ou tout ou partie de ces trois gènes, ainsi que les souches exprimant ce type de systèmes, notamment les souches de corynélactéries.

Dans les procédés mettant en oeuvre les constructions décrites, l'expression/sécrétion de l'amino-acide, polypeptide ou protéine déterminé sera régulée par la température, le milieu de culture et/ou la nature des sucres pour PS1 et PS2 et la concentration en sels (NH_4^+ notamment), métabolites (glutamate) et sucres (glucose/fructose) pour les systèmes à gdhA.

La présente invention concerne également les protéines comportant tout ou partie de la séquence de PS1 ou PS2, en particulier comportant un ou plusieurs sites antigéniques de ces protéines. Lesdites protéines peuvent être utilisées à titre d'élément atypique, notamment dans des trousse de diagnostic, de même que les anticorps correspondants.

L'invention concerne également les souches de corynélactéries dans lesquelles la protéine déterminée est ancrée sur la paroi par la partie de PS1 ou PS2 remplissant cette fonction d'ancre ou bien dans lesquelles les épitopes antigéniques de PS1 ou PS2 sont exposés sur la paroi.

Les exemples ci-après sont destinés à mettre en évidence d'autres caractéristiques et avantages de la présente invention et ne sont en aucun cas limitatifs.

Figure 1. Schéma du plasmide pCGL612. Plasmide dérivé du pUN121 (Nilsson, B., Uhlén, M., Josephson, S., Gatenbeck, S., and Philipson, L. (1983) An improved positive selection plasmid vector constructed by oligonucleotide mediated mutagenesis. Nucleic Acids Res 11: 8019-8029.) contenant un fragment de 2,6 kb de *C. melassecola* ATCC17965 portant l' intégralité du gène *csp1* permettant la synthèse de la protéine PS1.

5 Figure 2. Séquence nucléotidique et séquence d' acides aminés correspondante du gène *csp1* de *Corynebacterium glutamicum* dite *Corynebacterium melassecola* ATCC17965. La numérotation des nucléotides est présente sur le côté droit de la figure. Les séquences nucléotidiqes répétées sont entourées. La séquence probable Shine-Dalgarno est soulignée. Un palindrome de 24 pb qui correspond vraisemblablement au terminateur de transcription est indiqué par des flèches inversées. Cette séquence apparaît dans la banque de données de séquences nucléotidiqes EMBL sous le numéro d' accès X66078.

10 15 Figure 3. Carte de restriction de la région d' ADN de *C. melassecola* ATCC17965 séquencée portant *csp1*.

Figure 4. Alignement des séquences de la protéine PS1 de *C. glutamicum* et des protéines du complexe antigénique 85 de *Mycobacterium*. 85B *M. k.* représente l' antigène 85-B de *M. kansaii* (MIPSG16235). 85B *M. b.* représente l' antigène 85-B de *M. bovis* (MIPSC83179). 85B *M. l.* représente l' antigène 85-B de *M. leprae* (EMBLX60934). 85C *M. t.* représente l' antigène 85-C de *M. tuberculosis* (EMBLX57229). 85A *M. b.* représente l' antigène 85-A de *M. bovis* (MIPSA28544). 85A *M. t.* représente l' antigène 85-A de *M. tuberculosis* (MIPSI60062). Les séquences ont été alignées en utilisant le programme FastA de "Genetics Computer Group" (University of Wisconsin, USA). La numérotation des résidus est donnée pour chaque protéine au début de chaque ligne. Les résidus d' acides aminés similaires trouvés entre les différentes protéines sont entourés. Les résidus considérés comme similaires sont les suivants; acides ou amides (D, E, N, Q); basiques (H, K, R); polaires (P, A, G, S, T); non polaires (I, L, M, V) et aromatiques (F, W, Y). Les résidus d' acides aminés identiques entre les sept protéines sont indiqués par une étoile au-dessus des résidus concernés.

30 remarque: pour chaque antigène, le numéro d' accession est précisé associé au nom de la banque de données considérée et figurent entre parenthèses.

Figure 5. Interruption du gène *csp1*.

Intégration dans le chromosome d' un gène *csp1* interrompu. pCGL613' est non répliquatif chez *C. glutamicum*, il contient *csp1* (zone noircie) interrompu par le gène *aphA3* (Km). wt, *B. lactofermentum* 15 sauvage et $\Delta csp1$, l'intégrant contenant *csp1* interrompu.

5

Figure 6. Construction du plasmide pCGL616.

Le plasmide pCGL616 correspond au plasmide pCGL125 doté du gène *csp1* de *C. glutamicum*.

Figure 7. Plasmides permettant la synthèse de protéines PS1 tronquées.

10 Schéma des vecteurs dérivés de pCGL616, précision de la taille des protéines attendues et de leur détection (+) ou non (-) par Western blotting avec des anticorps polyclonaux anti-PS1.

15 **Figure 8.** Construction du plasmide pCGL1030. La région nommée A dans le schéma contient le promoteur de *csp1* suivi de la région d' ADN correspondant à la séquence signal de PS1 et des 30 premiers acides aminés de sa séquence mature.

20 **Figure 9.** Construction du plasmide 1031. La région nommée A est décrite dans la figure 8. La région de jonction entre PS1 et EGA a été séquencée et le détail de cette séquence est présenté.

25 **Figure 10.** Construction du plasmide 1032. La région nommée A est décrite dans la figure 8. La région de jonction entre PS1, (AQK)10 et EGA a été séquencée et le détail de cette séquence est présenté.

30 **Figure 11.** Construction du plasmide 1033. La région nommée A est décrite dans la figure 8. La région de jonction entre PS1, (AQ)19 et EGA a été séquencée et le détail de cette séquence est présenté.

35 **Figure 12.** Séquence nucléotidique et séquence d' acides aminés correspondante du gène *csp2* de *Corynebacterium glutamicum* dite *Corynebacterium melassecola* ATCC17965. La numérotation des nucléotides est présentée sur le côté droit de la figure. La séquence probable Shine-Dalgarno est soulignée. Un palindrome de 22 pb qui correspond vraisemblablement au terminateur de transcription est indiqué par des flèches inversées.

Figure 13. Carte de restriction de la région d' ADN de *C. melassecola* ATCC17965 séquencée portant *csp2*.

Figure 14. Interruption du gène *csp2* chez *C. glutamicum*.
Intégration chromosomique du gène interrompu. Le plasmide pCGL830 non répliquatif chez *C. glutamicum*, porte le gène *csp2* interrompu par le gène *aphIII*. Le sens de

transcription des gènes *aphIII* et *csp2* est représenté par une flèche sur le plasmide pCGL830. Wt, représente la souche *B. lactofermentum* 15 et *csp2 :: aphIII*, l'intégrant avec le gène *csp2* interrompu.

Figure 15. Translocation de PS1 en fonction de la température. 10 ml de culture en phase exponentielle ($D_0650=1$) à 34°C ont été marqués avec la méthionine ^{35}S (37 TBq/mmol, 16nM en concentration finale) durant 1 mn. A la fin du pulse du chloramphénicol (100 $\mu\text{g}/\text{ml}$) et de la méthionine ^{32}S (concentration finale 0,5 mM) sont alors ajoutés. 1ml d' aliquote est prélevé et rapidement refroidi à la température indiquée. L'incubation est continuée à cette température pendant 30 mn et la fraction sécrétée pariétale de PS1 est extraite. L' extrait est alors soumis à un SDS-PAGE et autoradiographié (a). L' intensité des bandes est déterminée par densitométrie (b, axe de gauche) et est donnée en unité arbitraire, sur une base de 100 à 34°C. La translocation est fonction de la transition de phase des lipides de la membrane.

Figure 16. Carte de restriction du gène *gdhA*.

Figure 17. Séquence complète du fragment NheI-BglI contenant le gène *gdhA* de *C. melassecola*.

Figure 18. Construction de pCGL141 et pCGL142, vecteurs de fusion entre le promoteur du gène *gdhA* et le gène *lacZ*.

Figure 19. Oligonucléotides utilisés dans les constructions.

Figure 20. Construction de pPROK(AQ)₂₀celA

Figure 21. Detail de construction plaçant le gène synthétique entre *ptac* et le gène *celA* - a) construction plaçant *celA* sous le contrôle de *ptac*, b) devenir de la construction après introduction du polypeptide AQ.

ptac : promoteur tac

25 RBS : site de liaison au ribosome

■ : séquence synthétique en 5 du gène *celA* permettant l'introduction de la séquence équivalente au polypeptide AG et sa fusion à d'autres gènes éventuels (DGF1 DGF2)

■ : nucléotides introduits au site BstX1 équivalents au polypeptide AQ

□ : séquence d'ADN équivalent à une partie de la séquence signal d'EGA

30 ■ : séquence d'ADN équivalent à la séquence codante d'EGA

— : terminateur de transcription

P : premier acide aminé appartenant à la séquence signal d'EGA

Figure 22. Structure de pCGL125.

Exemple 1. Identification de PS1 et de PS2 dans le surnageant de culture et dans la paroi de *Corynebacterium glutamicum*.

L'analyse d'un gel de polyacrylamide en condition dénaturante (SDS-PAGE) (Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.) d'un surnageant de culture de la souche *Corynebacterium melassecola* ATCC17965, actuellement redéfinie comme une souche de *Corynebacterium glutamicum* (Jones, D., and Collins, M. D. (1986) Irregular nonsporing Gram-positive rods. In *Bergey's Manual of Systematic Bacteriology*. Williams and Wilkins (eds). Baltimore, vol. 2, pp. 1261-1434.) montre deux protéines majeures nommées PS1 et PS2 de masse moléculaire approximative 67000 et 63000 respectivement. Les concentrations de PS1 et de PS2 suivent la courbe de croissance de la bactérie et atteignent leur maximum en phase stationnaire. Une fraction sécrétée importante de ces protéines et surtout pour PS2 est également située dans la paroi de la bactérie. Pour extraire PS1 et PS2 de la paroi un traitement des bactéries au SDS est utilisé, qui ne provoque pas de lyse significative de la bactérie. Ainsi, pour obtenir une concentration maximale de PS1 et de PS2, il est donc possible de cumuler les deux fractions sécrétées, surnageant de culture et fraction pariétale, et d'obtenir une préparation finale où PS1 et PS2 sont fortement majoritaires. Des anticorps polyclonaux ont été préparés contre PS1 et PS2, il n'y a pas de réaction immunologique croisée entre les deux protéines ce qui montre bien que ces protéines sont différentes. Des protéines ayant de fortes réactions immunologiques croisées avec PS1 et PS2 ont été trouvées dans le surnageant de culture de souches bactériennes apparentées à *Corynebacterium melassecola* ATCC17965 comme la souche *Brevibacterium lacticfermentum* 15 (Bonnassie, S., Oreglia, J., Trautwetter, A., and Sicard, A. M. (1990) Isolation and characterization of a restriction and modification deficient mutant of *Brevibacterium lacticfermentum*. *FEMS Microbiology Letters* 72: 143-146.), *Brevibacterium lacticfermentum* ATCC21086 et *Brevibacterium flavum* ATCC 14067. PS1 et PS2 ont été testées pour plusieurs activités enzymatiques incluant l'activité invertase, pectinase, nucléase, collagénase, amylase, bactériocine, endoglucanase et protéase à large spectre. Aucune de ces activités enzymatiques n'a pu être associée à PS1 ou PS2.

Exemple 2. Evidence de la fonctionnalité du peptide signal de PS1 chez Escherichia coli (Figure 1).

Lorsque le gène *csp1* porté par le plasmide pCGL612 (Figure 1) est exprimé chez *E. coli* TG1, l'analyse par Western blotting d'un extrait brut de ce recombinant avec les anticorps anti-PS1 révèle la présence d'une protéine majeure qui a la même masse moléculaire que la protéine PS1 présente dans le surnageant de culture de *C. melassecola*. Une protéine mineure de masse moléculaire légèrement supérieure est également détectée. En fait la bande protéique majeure correspond à la forme mature de PS1 (sans séquence signal) et la bande protéique mineure à la forme précurseur de PS1 (avec la séquence signal).

En effet, dans une première expérience, la libération des protéines périplasmiques (enzymes sécrétées) de la souche recombinante *E. coli* TG1(pCGL612) par choc osmotique (Heppel, L. A. (1967) Selective release of enzymes from bacteria. Science 156: 1451-1455.) et la détection du contenu en protéines libérées par Western blotting avec des anticorps anti-PS1 ne révèle que la protéine majeure. L'activité isocitrate dehydrogenase (Shioi, I., and Ujigawa, K. (1978) Enzymes of the glutamate and aspartate synthetic pathways in a glutamate-producing bacterium, *Brevibacterium flavum*. J Biochem 84: 647-657.) de la souche a été mesurée en guise de contrôle de lyse; cette lyse a été estimée dans cette expérience à moins de 1%. Ceci amène à la conclusion que la bande protéique majeure correspond à la forme mature de PS1 et que la protéine est exportée à travers la membrane cytoplasmique d'*E. coli*.

Dans une deuxième expérience, l'extrait brut de la souche recombinante *E. coli* TG1(pCGL612) a été analysé par Western blotting avec les anticorps anti-PS1 avant et après addition de chloramphénicol pour inhiber la synthèse protéique. La bande mineure disparaît progressivement après l'inhibition de la synthèse protéique. Cette bande mineure de PS1 ne disparaît pas si on ajoute 5 minutes avant l'addition de chloramphénicol, le CCCP (carbonyl cyanide m-chlorophenylhydrazone), un protonophore qui dissipe la force proton-motrice à travers la membrane cytoplasmique. La disparition de la bande mineure de PS1 n'est donc pas le résultat d'une dégradation par les protéases. Sa disparition progressive après inhibition de la synthèse protéique et son absence du périplasme sont en accord avec l'hypothèse de la maturation de cette forme précurseur par une séquence signal peptidase située dans la membrane et sa translocation à travers la membrane cytoplasmique. Ce résultat montre également que chez *E. coli* la maturation de PS1 est dépendante de la force proton-motrice, *in vivo*.

Exemple 3. Séquence nucléotidique du gène *csp1* codant pour PS1

Le séquençage d'un fragment de 2547 paires de base contenant le gène codant pour PS1, nommé *csp1*, et de la région en amont a été réalisé. La séquence nucléotidique est présentée dans la figure 2 (séquence ID n° 1). La figure 3 représente la carte de restriction de cette région séquencée.

Par analyse informatique, une phase ouverte de lecture de 1971 paires de base a été identifiée correspondant à 657 acides aminés.

Les signaux putatifs de démarrage de traduction (GAGAAGGAAAATTCATG) et de démarrage de transcription (TACATA(-35) et TAAGAT(-10) ont été identifiés. La séquence AGAAGGA extraite du site de fixation aux ribosomes décrit ci-dessus est complémentaire (soulignage) de l' extrémité 3' de l' ARNr des bactéries de type Gram positive *Staphylococcus aureus* et *Streptomyces lividans* (5'-GAUCACCUCCUUUCUOH-3') (McLaughlin, J. R., Murray, C. L., and Rabinowitz, J. C. (1981) Unique features of the ribosome binding site sequence of the Gram positive *Staphylococcus aureus* β -lactamase gene. J. Biol Chem.256: 11283-11291.) (Bibb, M. J., and Cohen, S. N. (1982) Gene expression in *Streptomyces*: Construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. Mol Gen Genet 187: 265-277.). La région d' ADN en 5' précédent le codon de démarrage de la traduction contient deux séquences nucléotidiques AAAAGTTATCCACAG et ATTGAAAAA répétées chacune deux fois, de 28 à 42 et de 70 à 84 pour la première, puis de 100 à 108 et de 171 à 179 pour la deuxième. Ces deux séquences pourraient être impliquées dans la régulation de la transcription du gène *csp1*.

En ce qui concerne les signaux de sécrétion, une séquence à l' extrémité NH₂ de la protéine présente les caractéristiques d'une séquence signal de bactérie de type Gram-positive (Watson, M. E. E. (1984) Compilation of published signal sequences. Nucleic Acids Res.12: 5145-5164.). Cette séquence signal comporte un excès de charge positive en position NH₂-terminale (7 acides aminés à charge positive dans les 18 premiers acides aminés), suivi d'une séquence avec un excès d'acides aminés non polaires (18 acides aminés non polaires dans les 23 acides aminés suivants) puis de deux séquences d'acides aminés putatifs d'un site de clivage de séquence signal (pro thr ala ile ala, en position 28 à 32) (pro met ala ser ala, en position 39 à 43). Parmi ces deux séquences d'acides aminés putatifs d'un site de clivage de séquences signal, la deuxième = pro met ala ser ala en position 39 à 43 semble la plus

probable; en effet, la protéine PS1 a été purifiée du surnageant de culture de *Corynebacterium glutamicum* jusqu'à homogénéité électrophorétique par deux protocoles différents (voir exemple 5) et les préparations ont été utilisées pour déterminer la séquence amino-terminale par la dégradation d' Edman. Aucun signal n'
5 a été obtenu, bien que 5 nmoles de protéine purifiée aient été utilisées. Comme deux protocoles de purification ont été utilisés, il est probable que la protéine PS1 soit bloquée *in vivo* et que le blocage ne soit pas une conséquence de la technique de purification utilisée. La deuxième séquence de clivage proposée ferait apparaître une glutamine (position 44) comme premier acide aminé de la séquence mature, facilement
10 convertie en acide pyroglutamique rendant impossible le séquençage amino-terminal de la protéine par la technique d' Edman.

Un site putatif de terminateur de type rho dépendant est trouvé dans la région 3' du gène à 55 nucléotides des trois codons stop och-amb-opa (Rosenberg, M., and Court, D. (1979) Regulatory sequences involved in the promotion and termination of
15 RNA transcription. Annu Rev Genet 13: 319-353.). Le ΔG de cette structure en épingle à cheveux est égal à -35,7 kcal/mole (Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83: 9373-9377.)

20 La masse moléculaire calculée correspondant aux 657 acides aminés contenus dans la phase de lecture ouverte est de 70874. Or la masse moléculaire de la séquence signal la plus probable (site de clivage entre l' acide aminé 42 et 43) est de 4411, ce qui donne une masse moléculaire calculée pour la protéine mature de 66463 et qui est très proche de la valeur de 67000 estimée sur gel de polyacrylamide dénaturant.

25 Les caractéristiques de la séquence sont rappelées ci-après:
de 239 à 244 TACATA (signal -35)
de 269 à 274 TAAGAT (signal -10)
de 405 à 414 GAGAAGGAAA site de fixation des ribosomes
de 420 à 2390 séquence codante
30 de 420 à 548 peptide signal de protéine secrétée
de 2455 à 2506 structure en épingle à cheveux, signal de terminateur de type rho dépendant .

Exemple 4. Homologies de séquences entre PS1 de *Corynebacterium glutamicum* et les protéines du complexe antigénique 85 de *Mycobacterium*. (Figure 4).

- 5 La moitié NH₂ de la protéine PS1 est très similaire aux trois antigènes mycobactériens sécrétés nommés 85-A, 85-B et 85-C (Closs, O., Harboe, M., Axelsen-Christensen, N. H., and Magnussen, M. (1980) The antigens of *Mycobacterium bovis*, strain BCG, studied by crossed immuno-electrophoresis: a reference system Scand J. Immunol 12: 249-263.) (Wiker, H. G., Harboe, M.,
- 10 Nagai, S., and Bennedsen, J. (1990) Quantitative and qualitative studies on the major extracellular antigen of *Mycobacterium tuberculosis* H37Rv and *Mycobacterium bovis* BCG. Am Rev Respir Dis 141: 830-838.). Les trois gènes correspondants de différentes espèces mycobactériennes ont été clonés et séquencés: antigène 85-A de *Mycobacterium bovis* BCG1173P2 et de *Mycobacterium tuberculosis* (Borremans, M., De Wit, L., Volckaert, G., Ooms, J., De Bruyn, J., Huygen, K., Van Vooren, J.-P., Stelandre, M., Verhofstadt, R., and Content, J. (1989) Cloning, sequence determination, and expression of a 32-kilodalton-protein gene of *Mycobacterium tuberculosis*. Infect Immun 57: 3123-3130.) (De Wit, L., De la Cuvellerie, A., Ooms, J., and Content, J. (1990) Nucleotide sequence of the 32 kDa-protein gene
- 15 (antigen 85A) of *Mycobacterium bovis* BCG. Nucleic Acids Res 18: 3995.), l' antigène 85-B de *Mycobacterium bovis* Tokyo, *Mycobacterium kansaii* et *Mycobacterium leprae* (Matsuo, K., Yamaguchi, R., Yamazaki, A., Tasaka, H., and Yamada, T. (1988) Cloning and expression of the *Mycobacterium bovis* BCG gene for extracellular a antigen. J. Bacteriol 170: 3847-3854.) (Matsuo, K., Yamaguchi, R., Yamazaki, A., Tasaka, H., Terasaka, K., and Yamada, T. (1990) Cloning and expression of the gene for the cross-reactive a antigen of *Mycobacterium kansaii*. Infect Immun 58: 550-556.) (De Mendonça Lima, L., Content, J., Van Heuverswyn, H., and Degrave, W. (1991) Nucleotide sequence of the gene coding for the 85-B antigen of *Mycobacterium leprae*. Nucleic Acids Res 19: 5789.), et l' antigène 85-C
- 20 de *Mycobacterium tuberculosis* (Content, J., De La Cuvellerie, A., De Wit, L., Vincent-Levy-Frébault, V., Ooms, J., and De Bruyn, J. (1991) The genes coding for the antigen 85 complexes of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG are members of a gene family: cloning, sequence determination, and genomic organization of the gene coding for antigen 85-C of M. tuberculosis. Infect Immun 59: 3205-3212.). La protéine PS1 de *Corynebacterium glutamicum* présente environ 33%
- 25
- 30
- 35

- de résidus identiques ($\sigma=1,1$) et environ 52% de résidus similaires ($\sigma=1,1$) avec ces six protéines sur une longueur d'environ 330 acides aminés (+/- 5). Cette longueur d'environ 330 acides aminés correspond pour les antigènes mycobactériens à la longueur totale de la protéine. Toutes ces protéines mycobactériennes, tout comme PS1, contiennent une séquence signal d'une longueur comparable aux plus longues séquences signal trouvées chez les bactéries de type Gram-positives (environ 42 acides aminés, $\sigma=2,4$). La protéine 85-B de *M. bovis* et la protéine 85-C de *M. tuberculosis*, tout comme PS1, ont une plus longue région NH₂ hydrophile (5 ou plus, de résidus chargés positivement) que la plupart des séquences signal. Une autre 5 caractéristique intéressante de toutes ces séquences signal est la présence en position 3 ou 5 d'un résidu acide, l'acide aspartique, sauf pour l'antigène 85-C de *M. tuberculosis* où il s'agit d'un acide glutamique. La présence d'un résidu chargé acide est commun aux extrémités NH₂ des séquences signal Eucaryotes, mais est tout à fait inhabituelle aux extrémités NH₂ des séquences signal Prokaryotes (Perlman, D., and 10 Halvorson, H. O. (1983) A putative signal peptidase recognition site and sequence in eucaryotic and prokaryotic signal peptides. J Mol Biol 167: 391-409.) (Watson, M. E. E. (1984) Compilation of published signal sequences. Nucleic Acids Res. 12: 5145-5164.). La raison de cette particularité n'est pas connue. Aucune autre similitude 15 significative n'a été trouvée entre PS1 et d'autres protéines présentes dans les banques de données EMBL/ MIPS.
- 20

Exemple 5. Protocoles de purification de PS1 et PS2 utilisés en vue de la détermination de la séquence N-terminale.

25 **Protocole 1 :**

Les protéines PS1 et PS2 ont été purifiées à partir du surnageant de culture de *C. glutamicum* ATCC 17965 par électrophorèse préparative sur gel de polyacrylamide et électroélution.

Les bactéries, cultivées dans 200 ml de milieu riche LB à 34°C, sont récoltées 30 en phase stationnaire de croissance par centrifugation à 8000 g pendant 15 minutes et à 4°C. Les protéines du surnageant de culture sont ensuite précipitées au sulfate d'ammonium 60% et récoltées par centrifugation à 13000 g pendant 15 minutes à 4°C. Le culot est solubilisé dans 4 ml de tampon Tris HC1 10 mM pH 6.8 et la solution est ensuite dialysée pendant 24 heures à 4°C dans ce même tampon.

- L'extrait protéique dialysé obtenu après précipitation au sulfate d'ammonium est déposé sur un gel d'électrophorèse de format 16 x 20 x 0,75 cm. L'électrophorèse est réalisé selon le protocole décrit par Laemmli (1970) en utilisant un gel de concentration à 4% et un gel de séparation à 7,5%. La migration se fait en quinze heures à 40mA.
- 5 Les gels sont ensuite colorés au chlorure de cuivre selon le protocole décrit par Lee et al (1987). Les bandes protéiques correspondant aux protéines PS1 et PS2 sont découpées puis totalement décolorées. Les protéines sont ensuite électroélues du gel pendant 5 heures à 48mA et à 4°C, puis dialysées plusieurs fois dans un tampon Tris HC1 10 mM pH 6.8 avant d'être réparties en plusieurs parties aliquotes et congelées à - 10 20°C. Le rendement de la purification est de l'ordre de 25% avec une pureté supérieure à 90%.

Protocole 2 :

- Les protéines PS1 et PS2 sont purifiées à partir du surnageant de culture de *C. glutamicum* ATCC 17965 par ultrafiltration, électrophorèse et transfert sur membrane de PVDF.

Les bactéries, cultivées dans le milieu riche LB à 34°C, sont récoltées en phase stationnaire de croissance par centrifugation à 8000 g pendant 15 minutes et à 4°C. 4 ml de surnageant sont dilués 50 fois dans un tampon phosphate 50 mM pH 7.0 avant 20 d'être centrifugés sur une membrane d'ultrafiltration dont le seuil de coupure est de 30 kD. Cette étape permet d'obtenir un extrait protéique de 80µl qui est ensuite déposé sur un gel d'électrophorèse composé d'un gel de concentration à 4% et d'un gel de séparation à 7,5%. L'électrophorèse est réalisée selon le protocole décrit par Laemmli avec les modifications suivantes. Toutes les solutions servant à la préparation des gels 25 ainsi que le tampon de migration sont dégazés et contiennent 0,1 M de thioglycolate. De plus, le gel de séparation est soumis à un "pré-run" avant son utilisation. Toutes ces précautions sont prises dans le but d'éviter au maximum la formation de radicaux libres qui pourraient conduire à des modifications de l'extrémité N-terminale des protéines et par conséquent à un éventuel blocage de cette extrémité. A l'issue de 30 l'électrophorèse, les protéines sont transférées sur une membrane de PVDF. Cette étape se fait dans un tampon 50 mM Tris, 50 mM borate pH 8.0 pendant 60 minutes et à 50V. La membrane est ensuite colorée à l'amidoblack, ce qui permet de localiser et découper les bandes correspondant aux protéines PS1 et PS2. Les bandes protéiques sont ensuite décolorées et utilisées telles quelles pour le séquençage N-terminal.

(Laemmli, U. K. 1970. Cleavage of structure proteins during assembly of the head of bacteriophage T4. *Nature*, 227 : 680-685. - Lee, C. Levin, A. Branton, D. 1987. Copper staining : a five minute protein stain for sodium dodecyl sulfate polyacrylamide gels. *Anal. Biochem.*, 166 : 308-312.)

5

Exemple 6. Obtention de souche de *Corynebacterium glutamicum* ne synthétisant plus PS1, dite PS1-. (Figure 5).

La souche *C. glutamicum* dite *Brevibacterium lactofermentum* 15 est permissive à l' ADN modifié d' *E. coli* K12 (Bonnassie, S., Oreglia, J., Trautwetter, A., and Sicard, A. M. (1990) Isolation and characterization of a restriction and modification deficient mutant of *Brevibacterium lactofermentum*. *FEMS Microbiol Letters* 72: 143-146), alors que la souche *C. glutamicum* dite *C. melassecola* ATCC17965 est une souche très restrictive vis à vis de l' ADN d' *E. coli* (Reyes O., Guyonvarch, A., Bonamy, C., Salti, V., David, F., and Leblon, G. (1991) 'Integron'-bearing vectors: a method suitable for stable chromosomal integration in highly restrictive *Corynebacteria*. *Gene* 107: 61-68.). Pour cette raison, la souche *B. lactofermentum* 15 a été choisie pour effectuer l' interruption du gène *csp1*. Il a été vérifié que la carte physique du gène *csp1* est identique chez *C. melassecola* ATCC17965 et chez *B. lactofermentum* 15.

Le fragment *Clal* de 1,5 kb du plasmide pAT21 (Trieu-Cuot, P., and Courvalin, P. (1983) Nucleotide sequence of the *Streptococcus faecalis* plasmid gene encoding the 3'5"-aminoglycoside phosphotransferase type III. *Gene* 23: 331-341.) contenant le gène *aphA3* de *Streptococcus faecalis*, qui confère la résistance à la kanamycine (Km^r), a été inséré dans le site unique *KpnI* (*Asp718*) du gène *csp1* présent dans le plasmide pCGL612, pour donner le plasmide pCGL613'. Par Western blotting avec des anticorps polyclonaux anti-PS1, il a été montré que la souche recombinante *E. coli* hébergeant le plasmide pCGL613' est bien de phénotype PS1-. Ce plasmide est capable de se répliquer chez *E. coli* mais pas chez *C. glutamicum*. Il a été introduit dans la souche de *C. glutamicum* dite *B. lactofermentum* 15 par électro-transformation (Bonamy, C., Guyonvarch, A., Reyes, O., David, F., and Leblon, G. (1990) Interspecies electro-transformation in *Corynebacteria*. *FEMS Microbiol Letters* 66: 263-270.) et les transformants Km^r ont été sélectionnés. Dans les transformants Km^r , le plasmide pCGL613' est supposé s' être intégré dans le chromosome de *C. glutamicum* par recombinaison homologue avec la région *csp1* du génome de l' hôte.

Dans 22,5% des transformants, un évènement de double crossing-over s'est produit, résultant dans la substitution du gène *csp1* sauvage par la construction *csp1::aphA3* du plasmide transformant, donnant un phénotype Km^r-Tet^s (Figure 5). L'ADN total chromosomal de la souche sauvage et d'un des recombinants Km^r-Tet^s digéré par soit 5 *Bg*II soit *Bam*HI et *Eco*RI a été analysé par Southern blotting (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning: a Laboratory manual, second edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Publications.) avec la sonde pCGL613'. Le gène *csp1* est contenu dans un fragment d'approximativement 7,5 kb dans la souche sauvage, alors que l'intégrant 10 pCGL613' contient un fragment d'approximativement 9 kb correspondant au gène *aphA3* de 1,5 kb inséré dans le gène *csp1*. La digestion *Bam*HI-*Eco*RI confirme la structure de l'intégrant présentée dans la figure 5.

L'intégrant Km^r-Tet^s a également été analysé par Western blotting pour la production de PS1 en utilisant des anticorps polyclonaux anti-PS1. Il n'y a pas de protéine PS1, ni dans le surnageant de culture, ni dans l'extrait brut de cette souche. Ceci confirme que le gène *csp1* cloné dans λgt11 correspond à un gène unique qui code bien pour PS1 dans *C. glutamicum*.

Cette souche *C. glutamicum* PS1- est tout à fait viable et ne semble pas affectée dans son taux de croissance. Ce résultat montre qu'il est possible d'utiliser la 20 région du gène *csp1* comme cible d'intégration d'ADN homologue ou hétérologue dans une souche de *C. glutamicum*, sans en affecter, a priori, la viabilité.

Exemple 7. Expression chez *C. glutamicum* du gène *csp1* en multicopies. Analyse des régions importantes de PS1, nécessaires à sa synthèse et à sa sécrétion.

5 Pour cette série d' expériences, le plasmide pCGL616 a été construit. Il contient l' intégralité du gène *csp1* et a été construit à partir du plasmide pCGL125, correspondant au plasmide pBL1 (Santamaria, R., Gil, J. A., Mesas, J. M. and J. F. Martin (1984) Characterization of an endogenous plasmid and development of cloning vectors and a transformation system in *Brevibacterium lactofermentum*. J. Gen. 10 Microbiol. 130: 2237-2246.), réplicatif chez *C. glutamicum*, doté d' une cassette de clonage comportant le gène *aphA3* de *Streptococcus faecalis* et du plasmide pCSP1G, réplicatif chez *E. coli*, contenant le gène *csp1*. Le plasmide, pCGL616, résultant de cette construction (Figure 6) est réplicatif chez *C. glutamicum*.

15 Restauration de la synthèse de PS1 dans une souche de *C. glutamicum* PS1-.

On observe la restauration dans la souche dite, *B. lactofermentum* 15 PS1-, de la synthèse de PS1, après introduction dans celle-ci du plasmide pCGL616. Dans cette souche *B. lactofermentum* 15 PS1- hébergeant le plasmide pCGL616, une quantité 20 plus importante de PS1 sécrétée est détectée en comparaison de la souche sauvage *B. lactofermentum* 15 (naturellement PS1+). Ceci signifie qu' il est possible d' augmenter la concentration de PS1 sécrétée dans une souche de *C. glutamicum* en augmentant le nombre de copies du gène *csp1*.

Ce résultat est également démontré dans la souche dite *C. melassecola* 25 ATCC17965.

Constructions de plasmides dérivés de pCGL616 permettant la synthèse de protéines PS1 tronquées. (Figure 7).

Cette expérience montre qu' une protéine PS1 tronquée d' une masse 30 moléculaire alors égale à environ 23000 (Mw), au lieu de 67000 (Mw) pour la protéine naturelle, peut encore être sécrétée chez *C. glutamicum*.

Sept délétions ont été effectuées dans la région du gène *csp1* donnant naissance à sept plasmides différents, à partir du plasmide pCGL616. Toutes ces délétions conservent la région de l' ADN équivalente à la séquence signal de PS1 ainsi 35 que le terminateur de transcription du gène *csp1*. Dans tous les cas la synthèse et la

sécrétion de la protéine PS1 tronquée a été analysée par Western blotting avec les anticorps polyclonaux anti-PS1. Ces résultats montrent qu'il est possible de déléter une grande partie du gène *csp1* (figure 7) tout en rendant encore possible la synthèse et la sécrétion d'une protéine PS1 tronquée. La plus grande délétion permettant ce 5 résultat correspond à la délétion du fragment *NcoI-BspEI* (*BspMII*) d'environ 1,3 kb du gène *csp1* (pCGL1041) donnant une taille de protéine précurseur pour PS1 égale à environ 29 kD et à environ 24kD pour la forme mature sécrétée. Dans le surnageant de culture de la souche dite *B. lactofermentum* 15 PS1- hébergeant ce plasmide pCGL1041, il est bien détecté par Western blotting avec les anticorps anti-PS1 une 10 protéine d'environ 23 kD.

Exemple 8. Construction d'un vecteur d'expression et de sécrétion chez *C. glutamicum* nommé pCGL1030 basé sur le système *csp1*: (Figure 8, 9, 10, 11)

15

Construction du plasmide pCGL1030 (Figure 8).

Ce plasmide réplicatif chez *C. glutamicum* (contient le plasmide pBL1 de *C. glutamicum*) porte le promoteur du gène *csp1* de *C. glutamicum* et la région d'ADN de ce gène correspondant à la séquence signal plus les 30 premiers acides aminés de la 20 séquence mature de PS1. Un site mutiple de clonage (polylinker 2 dans la figure 8) a été placé immédiatement derrière le 30^{ème} acide aminé de la séquence mature de PS1, afin de permettre le clonage aisément en phase de tout gène hétérologue devant être exprimé chez *C. glutamicum*. Enfin, ce plasmide est doté des éléments de PS1 nécessaires à la sécrétion et correspond donc à un outil et d'expression et de sécrétion.

25

Expression du gène *celA* de *Clostridium thermocellum* chez *Corynebacterium glutamicum* et sécrétion de la protéine correspondante (Figure 9).

Le gène *celA* de *C. thermocellum*, (Cornet, P., Millet, J., Béguin, P. and J.-P. 30 Aubert (1983) Characterization of two *cel* (cellulose degradation) genes of *Clostridium thermocellum* coding for endoglucanases. Bio/Technology 1:589-594.) codant pour une endoglucanase nommée endoglucanase A ou EGA, a été cloné dans le vecteur pCGL1030 au site *Sma I*, donnant naissance au plasmide pCGL1031(Figure 9). Ce 35 gène *celA* provient du plasmide pCGL1008 où un site de restriction *BstXI* a été introduit artificiellement, très près du site de démarrage de traduction de la protéine

EGA, à des fins de constructions chimères (voir Figure 10, 11). La synthèse de la protéine EGA est facilement détectable grâce à un test coloré d'activité enzymatique sur boîte utilisant un substrat des endoglucanases, la carboxyméthylcellulose dite CMC (Cornet, P., Millet, J., Béguin, P. and J.-P. Aubert (1983) Characterization of two *cel* (cellulose degradation) genes of *Clostridium thermocellum* coding for endoglucanases. Bio/Technology 1:589-594.). Ce test CMC sera utilisé pour confirmer la synthèse de la protéine EGA de *C. thermocellum* chez *C. glutamicum*. Un test CMC d'activité sur boîte effectué sur des cellules entières ou sur le surnageant de culture, en milieu riche (LB -Luria Broth- ou BHI -Brain Heart Infusion-) met en évidence dans les deux cas, l'activité endoglucanase d'une souche de *C. glutamicum* dite *Brevibacterium lactofermentum* 15 hébergeant le plasmide pCGL1031. On notera une activité plus forte sur milieu LB + fructose, ou + glucose, indiquant un effet stimulant de ces deux sucres sur l'expression de *celA* sous contrôle du promoteur *cspI*. Ceci se confirme dans le zymogramme (Béguin, P. (1983) Detection of cellulase activity in polyacrylamide gels using Congo red stained agar replicas Anal. Biochem. 131: 333-336.), et dans le Western blotting réalisé sur les surnageants de culture avec des anticorps polyclonaux anti-EGA.

20 Utilisation du système *cspI* pour l'expression et la sécrétion du polypeptide synthétique (AQK)10 (Figure 10).

Un gène synthétique correspondant au polypeptide alanine-glutamine-lysine répété 10 fois a été synthétisé chimiquement et cloné au site *Bst*XI du plasmide pCGL1008, donnant naissance au plasmide pCGL1017. Le fragment *Eco*RI du plasmide pCGL1017 a été cloné au site *Sma*I du plasmide pCGL1030, situé en aval 25 du promoteur *cspI* de la séquence signal et des 30 premiers acides aminés de PS1 (et en amont du gène reporter *celA*), donnant naissance au plasmide pCGL1032 (Figure 10). La détection de la protéine chimère PS1-(AQK)10-EGA est effectuée comme décrit ci-dessus, par test CMC sur boîte par zymogramme ou par Western blotting.

30 Utilisation du système *cspI* pour l'expression et la sécrétion du polypeptide synthétique (AQ)19 (Figure 11).

Un gène synthétique correspondant au polypeptide alanine-glutamine répété 20 fois a été synthétisé chimiquement et cloné au site *Bst*XI du plasmide pCGL1008, donnant naissance au plasmide pCGL1002. Le fragment *Eco*RI du plasmide 35 pCGL1002 a été cloné au site *Sma*I du plasmide PCGL1030, situé en aval du

promoteur *csp1* de la séquence signal et des 30 premiers acides aminés de PS1 (et en amont du gène reporter *celA*), donnant naissance au plasmide pCGL1033 (Figure 11). La détection de la protéine chimère PS1-(AQ)19-EGA est effectuée comme décrit ci-dessus, par test CMC sur boîte, par zymogramme ou par Western blotting. La 5 séquence chez B. lactofermentum dans le plasmide pCGL1033 a fait apparaître la perte d'une séquence codante AQ (passage de AQ₂₀ à AQ₁₉ lors du clonage chez B. lactofermentum).

Cette série d' expériences montre que le promoteur du gène *csp1* permet l' 10 expression chez *C. glutamicum* du gène hétérologue *celA* de *Clostridium thermocellum* et des constructions chimères (AQK)10 -*celA* et (AQ)19 -*celA*. De plus, ces expériences montrent que les éléments de PS1, en l'occurrence, sa séquence signal suivie des 30 premiers acides aminés de sa séquence mature placés en amont des 15 gènes hétérologues, permettent la sécrétion des produits correspondants. L' effet du milieu de culture et de l' adjonction ou non de sucre dans ce milieu, en l'occurrence de glucose ou de fructose, a un effet sur la production du produit correspondant. En particulier, sous le contrôle du promoteur de *csp1* et chez *C. glutamicum*, la production d' EGA ou des protéines chimères (AQK)10-EGA ou (AQ)19-EGA, est meilleure sur milieu LB que sur milieu BHI, elle est fortement stimulée par le glucose 20 ou le fructose sur milieu LB. Le promoteur *csp1* de *C. glutamicum* semble plus fort que le promoteur naturel de *celA* de *C. thermocellum*; en effet la souche dite B. lactofermentum 15 hébergeant le plasmide pCGL602, qui contient le promoteur naturel de *celA* présente une activité endoglucanase nettement moins importante que cette même souche hébergeant le plasmide pCGL1031, où *celA* est sous le contrôle du 25 promoteur *csp1* de *C. glutamicum*.

L' expérience de Western blotting réalisée sur les surnageants de culture de différentes souches contenant pCGL1032 ou pCGL1033 montre que plusieurs bandes protéiques réagissent avec des anticorps polyclonaux anti-EGA. Ces différentes bandes sont spécifiques de l' endoglucanase EGA (absentes du témoin) et 30 correspondent vraisemblablement à des produits de dégradation de la protéine et des protéines chimères. Cependant des bandes de plus hautes masses moléculaires sont bien observées avec (AQK)10-EGA (pCGL1032) et (AQ)19-EGA (pCGL1033) et de façon cohérente (Mw (AQ)19-EGA > Mw (AQK)10-EGA).

Exemple 9. Séquence nucléotidique du gène *csp2* codant pour la protéine PS2 de *Corynebacterium glutamicum* (Figure 12, 13).

Le séquençage d'un fragment de 2702 paires de base contenant le gène codant pour PS2, nommé *csp2*, et de la région en amont a été réalisé. La séquence nucléotidique est présentée dans la figure 12 (séquence ID n° 2). La figure 13 représente la carte de restriction de cette région séquencée.

Par analyse informatique, une phase ouverte de lecture de 1532 paires de base a été identifiée correspondant à 510 acides aminés.

Une séquence de type Shine Dalgarno a été identifiée, AAGGAG, juste en amont du codon de démarrage de la traduction (-12 à -17).

A l'extrémité NH₂ de la protéine se trouve une séquence signal tout à fait banale de bactérie de type Gram-positive de 30 acides aminés. Une séquence d'acides aminés, ile pro ala phe ala, putative d'un site de clivage de séquence signal a été trouvée. La détermination de la séquence amino-terminale de la protéine par la technique de dégradation d'Edman, purifiée à partir du surnageant de culture de *Corynebacterium glutamicum*, n'a donné aucun signal bien que 5 nmoles de protéine purifiée ait été utilisée. Comme deux protocoles de purification ont été utilisés, il est probable que la protéine PS2 soit bloquée *in vivo*, tout comme PS1, et que le blocage ne soit pas une conséquence de la technique de purification utilisée. La séquence signal proposée de 30 acides aminés fait apparaître une glutamine (position 31) comme premier acide aminé de la séquence mature, facilement convertie en acide pyroglutamique rendant impossible le séquençage amino-terminal de la protéine par la technique d'Edman. Cette protéine PS2 possède les caractéristiques des protéines de paroi, comme son caractère très acide (pI=4,1), son absence de résidus cystéine et son très faible contenu en résidus méthionine (Sleytr, U. B. (1978) Regular arrays of macromolecules on bacterial cell walls: structure, chemistry, assembly, and function. Int. Rev. Cytol. 53: 1-64.) (Sleytr, U. B. and P. Messner (1983) Crystalline surface layers on bacteria. Ann. Rev. Microbiol. 37: 311-339.). Les analyses de microscopie électronique confirme que PS2 est bien une protéine de paroi capable de s'arranger en structure hexagonale ordonnée à la surface cellulaire.

Un site putatif de terminateur de type rho indépendant est trouvé dans la région 3' du gène à 76 nucléotides du codon stop.

Les caractéristiques de la séquence sont les suivants:

de 562 à 567: site de fixation aux ribosomes
de 579 à 2108: séquence codante
de 579 à 668: séquence signal de protéine sécrétée
de 2188 à 2233: structure en épingle à cheveux, signal putatif de terminaison de
transcription de type rho indépendant (présent à 76 nucléotides du codon stop).

Exemple 10. Obtention de souche de *Corynebacterium glutamicum* ne synthétisant plus PS2, dite PS2-. (Figure 14).

- 10 L'interruption du gène *csp2* a été réalisée chez *C. glutamicum* dite *B. lactofermentum* 15 à l'aide du vecteur pCGL830 (Figure 14), vecteur non réplicatif chez les corynèbactéries, et portant une copie du gène *csp2* inactivé par l'insertion du gène *aphIII* (clonage du gène *aphIII* au site unique *Nru* I de *csp2* porté par le plasmide pCGL811). Aucun signal de PS2 n'a été mis en évidence par détection immunologique avec des anticorps polyclonaux anti-PS2, sur des extraits cellulaires issus de la souche *E. coli* TG1 portant le plasmide pCGL830. Par électroporation de la souche *B. lactofermentum* 15 et sélection sur Km, des clones intégrés ont été sélectionnés. Parmi ces intégrants, les clones Tet^s ont été obtenus indiquant un événement de double crossing-over conduisant à la substitution du gène sauvage par le gène interrompu.
- 15 L'analyse par southern blot de l'ADN chromosomique digéré par *Xba* I et *Sac* I des clones Km^r Tet^s en utilisant la sonde pCGL811 montre respectivement un fragment à 4,2kb et 2,2kb au lieu de 2,7kb pour *Xba* I et de 0,7kb *Sac* I obtenu pour la souche sauvage, indiquant une augmentation de taille liée à la présence du gène *aphIII*.
- 20 L'absence de détection de PS2 par Western blotting, avec des anticorps polyclonaux anti-PS2, dans les différentes fractions confirme l'interruption du gène *csp2* dans *B. lactofermentum*. Cette souche PS2- est tout à fait viable et n'est affectée en rien dans sa croissance. De même que la région du chromosome de *C. glutamicum* portant le gène *csp1*, cette région d'ADN portant le gène *csp2*, peut être également utilisée comme cible d'intégration d'ADN étranger sans affecter la croissance de la bactérie.
- 25
- 30

Restauration du phénotype PS2+ dans la souche *B. lactofermentum* 15 PS2-.

Le fragment *Scal-FspI* de 2,3 kb contenant l'intégralité du gène *csp2* ainsi que la région d'ADN en amont, a été sous-cloné dans le plasmide pCGL824 et réintroduit

5 dans la souche *B. lactofermentum* 15 PS2-, permettant la restauration du phénotype PS2+. Il est à noter qu'une plus grande quantité de PS2 est obtenue lorsque le gène est en multicopies. Ces résultats montrent que la quantité de produit sécrété issu du gène *csp2* de *C. glutamicum* peut être modifiée en fonction du nombre de copies du gène.

10 Une analyse en microscopie électronique d'un échantillon des souches PS2+ et PS2- (obtenue par la technique énoncée ci-dessus) par cryofracture montre très clairement que la protéine PS2 est effectivement une protéine de paroi capable de s'arranger en structure hexagonale ordonnée à la surface cellulaire.

15 **Exemple 11. Effet de la température sur la sécrétion de PS1. (Figure 15).**

Les bactéries en phase exponentielle de croissance (34°C) ont été marquées par la méthionine ^{35}S durant 1 mn. Du chloramphénicol (100 $\mu\text{g}/\text{ml}$) et un excès de méthionine froide (^{32}S) sont alors ajoutés (temps 0). La température de la suspension cellulaire est alors rapidement portée à la température désirée et l'incubation est continuée à la dite température pendant 30 mn. La translocation de PS1 est déterminée par SDS-PAGE, autoradiographie et quantifiée par densitométrie (Figure 15). La translocation de PS1 est clairement dépendante de la température. Aucune translocation n'a lieu au-dessous de 10°C, elle augmente rapidement au-dessus de cette température pour atteindre un maximum autour de 30°C. La translocation est corrélée à une transition de phase des lipides (figure 15).

Exemple 12. Construction d'une banque d'ADN chromosomique de *Corynebacterium melassecola* ATCC 17965 et clonage du gène *gdhA*

5 L'ADN chromosomique de la souche de *C. melassecola* ATCC 17965 a été préparé suivant la méthode décrite par Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (Eds) ((1987) Current protocols in Molecular Biology. John Wiley and Sons, New York). Une digestion ménagée par l'endonucléase de restriction *Mbo* I (Boehringer)
10 a été réalisée sur 10 µg de cet ADN en suivant le protocole décrit par Maniatis T, Fritsch EF, Sambrook J ((1982) Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Les fragments d'ADN ont été séparés en fonction de leur taille sur gradient de sucre comme décrit par Ausubel et al. (1987). Les fragments
15 d'une taille comprise entre 6 et 15 kb ont été retenus pour la construction de la banque.

Le plasmide de clonage pUN121 (Nilsson B, Uhlen M, Josephson S, Gatenberg S, Philipson L (1983) An improved positive selection plasmid vector constructed by oligonucleotide mediated mutagenesis. Nucleic Acids Res 11 : 8019-8030) a été préparé par la méthode de Birnboim HC, Doly J ((1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7 : 1513-1523), à partir de la souche de *E. coli* GM 2929 disponible librement auprès du Dr. B. Bachmann. Le plasmide a été linéarisé par l'endonucléase de restriction *Bcl* I (Boehringer).

25 La banque a été construite par ligation avec la T4 DNA ligase (Boehringer) dans les conditions décrites par Ausubel et al. (1987), de 1 µg de plasmide pUN121 linéarisé par *Bcl* I et de 2 µg des fragments d'ADN de 6 à 15 kb décrits ci-dessus. Le mélange de ligation a été introduit dans la souche de *E. coli* DH5 par électroporation en suivant le protocole décrit par Dower WJ, Miller JF, Ragsdale CW ((1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16 : 6127-6145). Les clones de *E. coli* portant les plasmides recombinants ont été sélectionnés directement par leur capacité à croître sur milieu LB contenant 10 µg/ml de tétracycline. Les plasmides de la totalité des clones résistants à la tétracycline ont été préparés par la méthode de Birnboim et Doly (1979). L'ensemble de ces plasmides correspond à la banque d'ADN.

- La souche de *E.coli* CLR207 recA (Mattaj IW, McPherson MJ, Wooton JC (1982) Localization of a strongly conserved section of coding sequence in glutamate dehydrogenase genes. FEBS Letters 147 : 21-25), déficiente pour l'activité glutamate dehydrogénase a été transformée avec la banque d'ADN de *C. melassecola* ATCC 17965. Un clone transformant de *E.coli* CLR 207 recA capable de croître sur milieu minimum de sélection contenant 100 µg/ml d'ampicilline a été sélectionné. Ce clone est porteur d'un plasmide recombinant, pCGL310. L'activité glutamate deshydrogénase mesurée suivant la méthode de Meers JL, Tempest DW, Brown CM ((1970)
- 5 10 Glutamine (amide) : 2-oxoglutarate amino transferase oxido-reductase (NADP), an enzyme involved in the synthesis of glutamate by some bacteria. J Gen Microbiol 64 : 187-194), est restorée dans la souche de *E.coli* CLR207 recA porteuse du plasmide pCGL310. Différents sous-clonages ont permis dans un premier temps de raccourcir le fragment d'ADN de *C. melassecola* portant le gène gdhA complet à un fragment d'ADN de 3,8 kb délimité par les sites de restriction Eco RI et Xho I. Une carte de restriction précise de ce fragment Eco RI - Xho I est présentée Fig 16. Des sous-clonages supplémentaires ont permis de délimiter plus précisément le gène gdh A à un fragment Nhe I - Bgl I de 2,2 kb. Une hybridation ADN-ADN par la méthode de Southern EM ((1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98 : 503-517), a montré que le fragment d'ADN cloné est bien issu de la souche de *C. melassecola* ATCC 17965.
- 15 20 25 Détermination de la séquence nucléotidique du gène gdh A.
Afin de procéder à la détermination de la séquence nucléotidique du fragment d'ADN Eco RI - Xho I mentionné ci-devant, les sous-clonages suivants ont été effectués :(1) Eco RI - Bgl II dans le vecteur M13 mp18 (Norrrander J, Kempe T, Messing J (1983) Construction of improved M13
30 vectors using oligodeoxy-nucleotide directed mutagenesis. Nucleic Acids Res 26 : 101-106) coupé par Eco RI - Bam HI, (2) Xba I - Pst I dans le vecteur M13 mp18 coupé par Xba I - Pst I, (3) Xho I - Bgl II dans le vecteur M13 mp18 coupé par Sal I - Bam HI, (4) Eco RI - Pst I dans le vecteur M13 mp19 (Norrrander et al. , 1983) coupé par Eco RI - Pst I. De la sorte, la
35 séquence nucléotidique complète du fragment Eco RI - Xba I contenu dans

le fragment Eco RI - Xho I a pu être déterminée sur les deux brins par la méthode de Sanger F, Nicklen S, Coulson AR ((1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74 : 5463-5467). La séquence complète du fragment Nhe I - Bgl I contenant le gène gdh A est
5 présentée Fig 17 (séquence ID n° 3).

Analyse de la séquence nucléotidique du gène gdh A

L'analyse de la séquence nucléotidique du fragment Nhe I - Bgl I permet de mettre en évidence les éléments suivants :

a) promoteur (nucléotides 1 à 572)

10 Le promoteur du gène gdh A peut être caractérisé en ce qu'il comprend les éléments structurels suivants :

_ nucléotides 251 à 266

signal TGGTCATATCTGTGCG présentant une similitude avec la séquence TGG(Py)A(Pu)NNNNTTGCA caractéristique des promoteurs 15 reconnus par le facteur σ₆₀ (Merrick MJ (1983) Nitrogen control of the nif regulon in Klebsiella pneumoniae : involvement of the ntr A gene and analogies between ntr C and nif A. EMBO J 2 : 39-44) et régulés par l'ammonium.

_ nucléotides 437 à 442

20 signal TTCACA présentant une similitude avec la séquence TTGAC(Pu) caractéristique de la zone -35 des promoteurs de Streptomyces sp. (Strohl WR ((1992) Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res 20 : 961-974))

_ nucléotides 466 à 471

25 signal TAGGAT présentant une similitude avec la séquence TAG(Pu)(Pu)T caractéristique de la zone -10 des promoteurs de Streptomyces sp. (Strohl, 1992)

_ nucléotides 558 à 572

signal GGGAACGAGGAAATC présentant une similitude avec la séquence 30 AAAGGAGGTGATC de fixation du ribosome chez Streptomyces sp. (Strohl, 1992)

b) séquence codante (nucléotides 573 à 1913)

La phase de lecture s'étendant de la position 573 à 1913 correspond à celle de la glutamate deshydrogénase en raison des données suivantes :

— La protéine déduite de cette phase de lecture comporte 447 acides aminés, avec un poids moléculaire prédit de 48957 Daltons. Ce poids moléculaire est très proche de celui du polypeptide (48300 D) observé après électrophorèse en gel dénaturant d'une préparation de glutamate deshydrogénase de la souche de C. melassecola ATCC 17965.

5 — La structure primaire de la glutamate deshydrogénase déduite de la séquence nucléotidique du gène gdh A de C. melassecola présente de fortes similarités avec les structures primaires de glutamate deshydrogénases d'autres organismes (Teller JK, Smith RJ, McPherson MJ, Engel PC, Guest JR ((1992) The glutamate dehydrogenase gene of Clostridium symbiosum : cloning by polymerase chain reaction, sequence analysis and over-expression in Escherichia coli. Eur J Biochem 206 : 151-159).

10 — Les acides aminés mentionnés par Baker PJ, Britton KL, Engel PC, Farrants GW, Lilley KS, Rice DW, Stillman TJ ((1992) Subunit assembly and active site location in the structure of glutamate dehydrogenase. Proteins 12 : 75-86) comme étant indispensables à l'activité glutamate deshydrogénase sont présents dans la glutamate deshydrogénase de C. melassecola, et ce à des positions équivalentes à celles décrites par Baker et al. (1992).

15 — La structure secondaire de la glutamate deshydrogénase de C. melassecola, déduite de la séquence primaire mentionnée ci-dessus, présente de fortes similarités avec les structures secondaires de glutamate deshydrogénases d'autres organismes (Teller et al. , 1992).

20 25 c) terminateur (nucléotides 1937 à 1977)

Le terminateur du gène gdh A peut être caractérisé en ce qu'il comprend l' élément structurel suivant :

— séquence CCCTGATCCCGCGTTAAGGATCAGGG pouvant former une structure en épingle à cheveux riche en appariements GC avec un $\Delta G = -$ 30 35 13,6 kcal / mole, suivie de la séquence TTATTGATTTCTT riche en T. Une telle structure est caractéristique des terminateurs rho-indépendants (Rosenberg M, Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann Rev Genet 13 : 319-353).

Régulation de l'expression du gène gdh A de C. melassecola

La régulation de l'expression du gène gdh A de C. melassecola ATCC 17965 a été étudiée par mesure des variations de l'activité spécifique glutamate deshydrogénase en fonction de la nature du milieu dans lequel

- 5 cette souche a été cultivée. L'activité glutamate deshydrogénase a été mesurée par la méthode de Meers et al. (1970) à partir d'extraits acellulaires de C. melassecola obtenus par ultrasonication.

Les milieux de culture utilisés pour cette étude sont des milieux synthétiques dont la base est celle décrite par Liebl W, Klamer R, Schleifer KH (1989) (Requirement of chelating compounds for the growth of *Corynebacterium glutamicum* in synthetic media. Appl Microbiol Biotechnol 32 : 205-210). Les modifications suivantes ont été apportées :

- La source de carbone est soit du glucose à 11 g/l final (milieux 1 , 2 et 4) soit du fructose à 10 g/l (milieu 3).
- 15 — La concentration en ions NH4+ est de 125 mM dans les milieux 1 , 3 et 4. Elle est de 1,25 mM dans le milieu 2 (limitation en NH4+).
- Le milieu 4 contient 50 g/l final de L-glutamate.

Les activités spécifiques mesurées pour la glutamate deshydrogénase de la souche C. melassecola ATCC 17965 cultivée dans les 20 différents milieux décrits ci-dessus sont données dans le tableau ci-dessous. Les activités sont exprimées en micromoles de NADPH2 transformées par minute et par milligramme de protéines.

	milieu	milieu 1	milieu 2	milieu 3	milieu 4
25	Act. spécifique GdhA	4,4 +/- 0,3	23,2 +/- 1,1	18,2 +/- 1,8	2,8 +/- 0,2

Ce tableau permet de mettre en évidence les trois types suivants de régulation de l'expression du gène gdh A de C. melassecola ATCC 17965.

- répression de l'expression par le glutamate (facteur 1,57)
- 30 — répression de l'expression par l'excès d'ammonium (facteur 5,27)
- répression catabolique par le glucose (facteur 4,13 entre fructose et glucose). Il faut noter que dans le cas de la répression catabolique, les activités enzymatiques isocitrate deshydrogénase, aconitase, citrate synthase sont également touchées.

Construction d'un vecteur de fusion gdh A - lac Z

Afin de contrôler chez C. melassecola le caractère transcriptionnel de la régulation du gène gdh A par le glutamate, l'excès d'ammonium et le glucose, et de disposer d'un outil permettant une sélection simple de 5 mutants de C. melassecola non soumis à ces régulations, une construction a été réalisée entre le promoteur et le codon ATG d'initiation de la traduction du gène gdh A et l'opéron lac d' E. coli déleté au niveau du gène lacZ de ses cinq premiers acides aminés. Cette fusion a été réalisée comme suit :

- Isolement d'un fragment Eco RI - Bsp HI contenant le promoteur 10 du gène gdh A.
 - Conversion de l'extrémité Bsp HI en extrémité franche.
 - Clonage du fragment ainsi obtenu dans le vecteur pMC 1403 (Casadaban MJ, Chou J, Cohen SN (1980) In vitro gene fusions that join an enzymatically active β -galactosidase segment to amino-terminal fragments 15 of exogenous proteins : Escherichia coli plasmid vectors for the detection and cloning of translational initiation signals. J Bacteriol 143 : 971-980) linéarisé par Eco RI et Sma I, donnant naissance au plasmide pCGL 133.
- Isolement du fragment Nhe I - Sal I de pCGL 133 contenant la 20 fusion promoteur gdh A - opéron lac décrite ci-dessus et clonage dans le vecteur pCGL 241 (Reyes O, Guyonvach A, Bonamy C, Salti V, David F, Leblon G (1991) "Integron" bearing vectors : a method suitable for stable chromosomal integration in highly restrictive Corynebacteria. Gene 107 : 61-68) linéarisé par Spe I et Sal I, donnant ainsi naissance au plasmide pCGL 140 (figure 18).
- Transfert de l'intégron isolé de pCGL 140, contenant la fusion 25 gdhA - lac ainsi que le gène aph III conférant la résistance à la kanamycine, dans le vecteur pCGL 125, donnant naissance à pCGL 141 et pCGL 142 (Fig. 18). Les plasmides pCGL 141 et pCGL 142 ont été introduits dans la souche C. melassecola ATCC 17965 par transformation.
- La fonctionnalité de la fusion gdh A - lac a été montrée par mise en 30 évidence d'une activité β -galactosidase dans les souches de C. melassecola ATCC 17965 transformées par pCGL 141 et pCGL 142, activité absente de la même souche transformée par pCGL 125. L'activité β -galactosidase est mise en évidence par culture des bactéries sur milieu complet solidifié (BHI, 35 Difco) contenant le substrat chromogène X-gal (5-bromo-4-chloro-3-

indolyl β -D-galactopyranoside). Les colonies issues de bactéries possédant l'activité β -galactosidase deviennent bleues sur un tel milieu. Par culture des bactéries transformées par pCGL 141 et pCGL 142 sur les milieux 1, 2, 3 et 4 décrits ci-dessus, solidifiés par adjonction d'agar à 15 g/l final, et supplémentés par de la kanamycine à 25 mg/l final et du X-Gal à 100 mg/l final, nous avons pu montrer que la régulation du gène gdh A est bien de type transcriptionnel puisque les colonies bactériennes obtenues sur ces différents milieux présentent un gradient de coloration compatible avec la régulation montrée par mesure enzymatique. En effet, les colonies obtenues sur milieu 4, sont d'un bleu plus clair que celles obtenues , dans l'ordre d'intensité croissante, sur les milieux 1, 3 et 2.

Nous avons montré que cette différence se reflétait au niveau de la mesure enzymatique de l'activité β -galactosidase pour des cultures de C. melassecola transformée par pCGL 141 en milieu 1 et milieu 4 (répression par le glutamate).

milieu	milieu 1	milieu 4
Act sp. β -gal.	0,118	0,052

Les activités β -galactosidase ont été mesurées comme décrit par Miller JH (1972) (*Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), à partir d'extraits acellulaires de C. melassecola.

Sélection de mutants déficients en répression catabolique

Une mutagénèse NTG d'une souche dérivée de C. melassecola ATCC 17965 a été réalisée. A partir de cette mutagénèse, une première sélection a été appliquée sur le critère de résistance à un analogue du glutamate, le 4-fluoroglutamate. Les mutants résistants à cet analogue peuvent appartenir à différentes classes dont la classe de non répression catabolique. En effet, dans de tels mutants, on peut s'attendre à ce que l'élévation d'activité spécifique de la glutamate deshydrogénase conduise à une surproduction de glutamate intracellulaire, et ainsi à une dilution de l'analogue toxique, d'où le phénomène de résistance. Les mutants de

réistance au 4-fluoroglutamate ont été rassemblés et l'ensemble des cellules a été soumis à transformation par pCGL 141. Les bactéries transformées ont été étalées sur milieu 1 solidifié contenant du X-Gal et de la kanamycine. Les colonies bactériennes présentant la couleur bleue la plus intense ont été isolées, et mises en culture en milieu 1 liquide contenant de la kanamycine. L'activité glutamate deshydrogénase a été mesurée à partir d'un extrait acellulaire, l'activité β -galactosidase à partir de cellules entières toluénisées (Miller, 1972). Les résultats obtenus pour l'un des mutants sélectionnés sont présentés ci-dessous.

10

activité	glutamate deshydrogénase	β -galactosidase
témoin	6,3	10,79
mutant 90	12,1	23,66

15

Les résultats obtenus montrent donc bien qu'il est possible de sélectionner par cible phénotypique des mutants de régulation du gène gdh A avec l'outil construit. Il faut noter qu'il est très aisément d'éliminer pCGL 141 et pCGL 142 des cellules après sélection, simplement par culture en absence de pression de sélection kanamycine.

20

Exemple 13. Construction d'un plasmide permettant le clonage de peptides

Pour cette construction, une étape de sous-clonage de celA a été réalisée. Le gène celA disponible sous forme d'un fragment HindIII de 3,5kb contenant la région promotrice, le gène et le début d'un autre gène non identifié, a été sous-cloné sous forme d'un fragment HindIII-EcoRI de 2,6kb déléte du morceau de gène inconnu, dans un vecteur réplicatif d'E. coli, le pMTL23 (Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton N.P. (1988) The pMTL nic - cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene. 68: 139-149.)

Le site EcoRI a été introduit par mutagénèse dirigée immédiatement derrière le terminateur de transcription du gène. Ce sous-clonage intermédiaire, compte tenu des sites de restriction introduits est nécessaire à l'étape suivante de clonage ; en particulier le clonage dans le polylinker de pMTL23 permet l'introduction d'un site de restriction Ncol juste derrière le site EcoRI, ce qui permet de sortir le fragment contenant la région codante de celA sous forme d'un fragment NaeI-Ncol. Cette étape permet également de disposer d'un gène celA dépourvu de séquences non identifiées en 3'.

Le clonage de celA sous forme d'un fragment HindIII-EcoRI a été réalisé dans le plasmide pMTL23 en utilisant la souche réceptrice d'E. coli TG1. La souche d'E. coli possédant ce plasmide est bien dotée du phénotype CMC+ associé à l'expression de l'EGA ; l'analyse des fragments de restriction obtenus est conforme à ce qui est attendu.

La construction de pPROK-celA (figure 20) est la suivante : On utilise le plasmide pPROK-1 de 4,6kb disponible chez Clontech Laboratories, Inc. (Palo Alto, CA, USA).

Ce plasmide réplicatif chez E. coli contenant le promoteur tac (Brossius et coll. Gene 27:161, 1984) est hydrolysé par EcoRI-Ncol.

On introduit ensuite dans cette restriction les adaptateurs DGF1/DGF2 de la figure 19 sous forme EcoRI-Blunt, ces adaptateurs créent le site BstXI. Puis on introduit celA sous forme NaeI(blunt)-Ncol à partir de la construction précédente.

Le plasmide ainsi obtenu est dénommé pPROK-celA. Il importe le gène celA sous le contrôle du promoteur tac séparé par un site BstXI introduit grâce aux adaptateurs DGF1/DGF2.

Exemple 14. Construction d'un plasmide permettant l'expression de séquences AQ multiples

Pour réaliser l'insertion d'une séquence codant pour 20 unités Ala-Gln (AQ) on utilise un deuxième couple d'oligonucléotides de synthèse 5 dénommés DGF5/DGF6 (figure 19) qui correspondent au gène synthétique :

5' CAG[AQ]₂₀ CAGGCA 3'

3' CCGTGTC[AQ]₂₀ GT 5'

[AQ] représentant la séquence codant pour Ala-Gln.

Les extrémités des séquences de DGF5 et DGF6 sont 10 compatibles avec le site BstXI et la séquence peut donc être clonée dans ce site.

Les séquences d'extrémités de DGF5 et DGF6 sont telles que, d'une part, elles orientent la direction du clonage, et, d'autre part, elles détruisent le site BstXI à la suite du clonage.

15 L'utilisation d'adaptateurs non phosphorylés évite d'introduire plusieurs gènes synthétiques en tandem.

Après digestion du pPROK-celA (figure 20) par BstXI et ligation du gène synthétique, on obtient : pPROK(AQ)₂₀celA ayant la structure représentée figure 20.

20 La figure 21 détaille plus particulièrement la structure du site de fusion AQ/EGA et montre l'intérêt du site BstXI utilisé. Ce site est de structure :

CCATGGCAATGG

On constate qu'il comporte un codon de départ ATG ainsi que 25 le codon codant pour lalanine, GCA et un 2ème codon ATG pour linsertion d'une méthionine après la séquence codante déterminée.

L'insertion de l'adaptateur DGF5/DGF6 ne peut se faire que dans un sens et n'introduit aucune base étrangère à l'objet visé.

30 Ce plasmide est traité par BamHI et traité par ligation avec le produit de restriction du plasmide pCGL125 traité par la même enzyme. Le plasmide pCGL125 (figure 22) est un plasmide fonctionnel de Brevibacterium lactofermentum 15 comportant une origine de réPLICATION pBL1.

Par transformation de ladite souche par le plasmide (pCGL125-(AQ)₂₀-celA) (pCGL1002 figure 11) et sélection des souches transformées on obtient une souche selon la présente invention.

Dans toutes les fusions qui sont réalisées, la traduction commence par une méthionine immédiatement suivie par (AQ)₂₀ ; on a également pris la précaution de border (AQ)₂₀ par une méthionine en COOH-terminal ; la détection du polypeptide AQ fusionné ou non à la protéine celA peut se faire grâce à des anticorps spécifiques ou par détection analytique après purification sommaire de la protéine de fusion et hydrolyse au bromure de cyanogène ou inversement. Les propriétés particulières des peptides répétés permettent une séparation aisée.

Les souches citées ont les origines suivantes :

Escherichia coli

- | | | |
|----|---------------|------------------|
| 15 | . CLR207 recA | B. Bachman |
| | . DH5alpha | Gibco BRL |
| | . GM2929 | B. Bachman |
| | . TG1 | Institut Pasteur |

Brevibacterium flavum

- | | | |
|--|--------------|------|
| | . ATCC 14067 | ATCC |
|--|--------------|------|

20 Corynebacterium glutamicum (Brevibacterium lactofermentum)

- | | | |
|--|--------------|-------------|
| | . 15 | S. Bonassie |
| | . ATCC 21086 | ATCC |

Corynebacterium glutamicum (Corynebacterium melassecola)

- | | | |
|--|--------------|------|
| | . ATCC 17965 | ATCC |
|--|--------------|------|

25 La souche DH5alpha est disponible dans le catalogue de Clontech laboratories n° C1021-1 (Palo Alto, CA, USA).

Les souches ATCC sont disponibles à American Type Culture Collection c/o Sales and Marketing Departement, 12301 Parklawn Drive, Rockville, MD 20852 USA.

30 Une souche a été déposée dans la Collection Nationale de Cultures de Microorganismes (CNCM) de l'Institut Pasteur (Paris) le 23 juillet 1991 :

- . *Brevibacterium lactofermentum* 15 (CGL2005(B115) sous le n° I-1126.